MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 8: 359–379 (2024)
DOI: 10.3897/mbmg.8.125095

Emerging Technique

NAMERS: a purpose-built reference DNA sequence database to support applied eDNA metabarcoding

Kristen M. Westfall[1], Gregory A. C. Singer[2], Muneesh Kaushal[2], Scott R. Gilmore[1], Nicole Fahner[2], Mehrdad Hajibabaei[2,3], Cathryn L. Abbott[1]

1 *Fisheries and Oceans Canada, Pacific Biological Station, Nanaimo, BC, Canada*
2 *Centre for Environmental Genomics Applications/eDNA tech, St. John's, NL, Canada*
3 *University of Guelph, Guelph, ON, Canada*

Corresponding author: Kristen M. Westfall (westfall.kristenm@gmail.com)

Abstract

Applied eDNA metabarcoding is increasingly being considered as a tool to inform management decisions, regulations, or policy development. Because these downstream considerations are coming to the forefront of eDNA applications, optimizing workflow elements is essential to increasing standardization, efficiency, and competency of metabarcoding results. Reference DNA sequences are critical workflow elements that currently lack consistent approaches to generating, curating, or publishing. We present a complete mitochondrial genome and nuclear ribosomal DNA cistron reference DNA sequence library for 92% of the freshwater fish species of British Columbia, Canada. This resource is published as the Novel Applied eDNA Metabarcoding Reference Sequences (NAMERS) repository (https://namers.ca), a user-friendly and interactive website for specialists and non-specialists alike to explore and generate custom reference libraries for taxa and genes of interest. We demonstrate the power of NAMERS for optimization of applied eDNA metabarcoding study design by analyzing the number of primer mismatches and species resolution power of existing metabarcoding markers. NAMERS demonstrates that high quality curated genomic information is within a reasonable reach to meet the increasing demand for actionable eDNA metabarcoding applications. The framework used here incorporating the pillars of accuracy, completeness and accessibility can be applied for new iterations of other reference sequence databases to bring DNA-based monitoring into a new era.

Key words: Biodiversity, environmental DNA, fish, freshwater, species at risk, standardization

OPEN ACCESS

This article is part of:
Towards standardized molecular biodiversity monitoring
Edited by Teresita Porter, John Darling, Kelly Goodwin, Tiina Laamanen, Kristian Meissner, Toshifumi Minamoto

Academic editor: Tiina Laamanen
Received: 9 April 2024
Accepted: 13 October 2024
Published: 19 November 2024

Citation: Westfall KM, Singer GAC, Kaushal M, Gilmore SR, Fahner N, Hajibabaei M, Abbott CL (2024) NAMERS: a purpose-built reference DNA sequence database to support applied eDNA metabarcoding. Metabarcoding and Metagenomics 8: e125095. https://doi.org/10.3897/mbmg.8.125095

Copyright: This is an open access article distributed under the terms of the CC0 Public Domain Dedication.

Introduction

The need has never been greater for information rich biomonitoring data to assess environmental impacts, monitor rare species, chart ecosystem trajectories, and evaluate remediation and conservation efforts - all to ultimately maximize positive desired outcomes. There is great current attention on translating the science of eDNA metabarcoding into practices that benefit humankind (Cordier et al. 2021; Cristescu and Hebert 2018; Hering et al. 2018; Zaiko et al. 2018; Ruppert et al. 2019; Wang et al. 2021; Schenekar 2022). Despite the proven usefulness of

eDNA metabarcoding for biomonitoring (Kelly et al. 2024), widespread uptake for decision-making requires additional attention to workflow elements that are essential for translating the science to a broad end-user community (Stein et al. 2024). Specifically, the optimization of factors related to operational feasibility, cost, taxonomic breadth, throughput, and scalability is needed (as overviewed in other works: Baird and Hajibabaei 2012; Hering et al. 2018; Cordier et al. 2021), as there are many potential sources of biases in the metabarcoding workflow (Zinger et al. 2019). Indeed, the lack of quality criteria or standards and unknown confidence in results are serious impediments to the uptake of eDNA methods for applications (Darling and Mahon 2011; Stein et al. 2024).

Environmental DNA metabarcoding results that directly inform a management or regulatory decision require a higher level of confidence than results that are used in research and development to advance the science. Confidence in eDNA results is elevated by implementing strict quality criteria and standards throughout the workflow, and defining these quality criteria and standards becomes critical as eDNA gains popularity with government organizations with an eye to using this technology to inform policy, regulation, or management actions. Here we adopt the term applied eDNA metabarcoding to distinguish between end uses of eDNA results; the term applied meaning results is potentially used to inform management decisions, regulation, policy, etc. Just as Darling et al. (2020) argued for the need to distinguish between high throughput sequencing studies designed for ecological versus biosecurity surveillance purposes, there needs to be a similar distinction for eDNA metabarcoding.

DNA barcoding emerged as a powerful tool for genetic species identification before the naissance of eDNA metabarcoding almost a decade later (Hajibabaei et al. 2011; Taberlet et al. 2012). DNA barcoding uses a standardized DNA sequence proven to provide species-level resolution and a purpose-built, curated database (the Barcode of Life Data System, Ratnasingham and Hebert 2007) to determine the taxonomy of unknown specimens. In contrast, eDNA metabarcoding often uses multiple markers not proven to amplify all target taxa or to provide species level resolution therein, and assigns taxonomy to unknown sequences using data from massive online global DNA sequence repositories (Cordier et al. 2021; Wang et al. 2021) (e.g. GenBank, Benson et al. 2000). GenBank is by far the most popular repository for taxonomic assignment of environmental sequences. Although GenBank's RefSeq project (O'Leary et al. 2016) has improved the curation level of organellar genomes in GenBank, minimum specimen metadata or museum specimens are not required and not all GenBank records have been validated against a RefSeq sequence.

There are three foundational pillars of reference DNA sequence databases that underpin their use in applied eDNA metabarcoding to increase end user confidence in results; quality, completeness, and accessibility. The unique needs related to quality and completeness are false positives and false negatives (Leese et al. 2018; Zinger et al. 2019; Mathieu et al. 2020) that may arise from incorrect species identification, taxonomically incomplete data sets, or other factors such as unassessed genetic variation in widespread species. To address quality issues, using vouchered specimens from museum collections that have a perpetual link between specimen and sequence data elevates defensibility by matching an environmental sequence with a physical specimen for decades after the reference sequence database has been generated.

Furthermore, prioritizing a taxonomically complete library may reduce false positives when an environmental sequence matches more than one species. Genetically complete libraries prioritizing complete mitochondrial genomes enable data driven survey design by allowing for species resolution and primer mismatch to be evaluated at the outset when choosing fit-for-purpose markers appropriate to the survey objectives. Finally, accessibility is often an overlooked but essential component for reference DNA sequence databases. Curating sequences from large repositories takes a high level of skill and effort, and this step can introduce errors or be a stumbling block for non-specialist end users.

Here we present the development of a mitogenome (and ribosomal DNA cistron) reference sequence database and web portal that combines these three pillars into a single framework, representing a level of functionality that is not achieved by any current single existing repository. NAMERS: Novel Applied eDNA Metabarcoding Reference Sequence database, is a database of whole mitogenomes and nuclear ribosomal (nr) DNA cistrons for freshwater fish in British Columbia, Canada. The majority of sequences in NAMERS come from vouchered specimens with permanent records in public institutions that have an associated minimum of standard biodiversity metadata terms that include geo-referenced collection sites, dates, and taxonomic identification. NAMERS contains mitogenomes and nrDNA for 92% of all freshwater fish species in BC, thus offering high genetic and taxonomic completeness. Lastly, the NAMERS database is available in a user-friendly web portal that combines functionality with ease-of-use, requiring no bioinformatics experience. Users can easily view multiple sequence alignments for taxa and genes of their choice, and download customized reference DNA sequence data with a few clicks. This level of accessibility by specialists and non-specialist end users is not readily achieved by other repositories.

The NAMERS framework is based on the following as foundational premises of applied eDNA metabarcoding: (i) primer design is a key factor determining success (Piñol et al. 2015; Ficetola et al. 2021), and primers must be chosen carefully to meet application-specific needs (Taberlet et al. 2012; Deagle et al. 2014); (ii) achieving accurate species-level resolution is important, as higher taxonomic levels lack sufficient information for most biomonitoring needs (Baird and Hajibabaei 2012); (iii) different genes work best for different taxa (Creer et al. 2016; Ficetola et al. 2021), and multi-marker methods out-perform single marker ones if reference databases are sufficiently populated (McElroy et al. 2020); and (iv) mitochondrial markers are the workhorses of vertebrate eDNA metabarcoding with gene regions other than COI often working best (Collins et al. 2019; Deagle et al. 2014).

The framework presented here fills an innovation gap between the existing state of most reference DNA sequence databases and what is needed for managers and other end users when considering the downstream applications of eDNA metabarcoding results. This framework is currently presented as a proof-of-concept at the regional scale. Although there are immediate benefits for the management of freshwater fish in BC, the value of this framework goes well beyond this scale and we promote its use at the national and international levels. A framework like this has not thus far been implemented at larger scales due to the lack of organization and long-term funding. Environmental DNA is quickly gaining momentum with releases of the US's National Aquatic Environmental DNA Strategy (Goodwin et al. 2024), the EU's Marine Strategy Framework Directive

(European Parliament & Council of the European Union 2008), Australia's National Biodiversity DNA Library (https://research.csiro.au/dnalibrary/), and other global initiatives to increase the use of this technology across organizations and applications. The three pillars of reference DNA sequence databases outlined here and implemented in NAMERS are mission-critical to advancing the use of eDNA within the context of these and future major initiatives.

Methods

Fish sampling

To maintain consistently high quality of information in the NAMERS database, we aimed to satisfy three criteria for all species: (1) available museum-catalogued voucher specimen; (2) minimum voucher specimen metadata consisting of collection site name, geographic location, sampling date, and the name and affiliation of who did the morphological identification; and (3) genetic species identity verification using COI barcode sequences.

British Columbia (BC) has approximately 92 (75 native and 17 invasive) fish taxa that use freshwater for all or part of their life cycle, including significant geographical variants or subspecies (Beamish 1987; Taylor et al. 1999; McPhail 2007; Ruskey and Taylor 2016). Samples from at least one individual for 85 of these taxa were obtained, either from tissue samples (frozen or in ethanol) and/or DNA from curated voucher specimens. As a priority, voucher specimens were obtained from within their BC range (n = 35), and secondarily from other areas of Canada or the USA (Suppl. material 1). This included Washington (n = 4), Alaska (n = 2), Oregon (n = 1), Yukon (n = 9), Alberta (n = 3), Manitoba (n = 1), Ontario (n = 17), and Quebec (n = 12). Details on voucher specimens are in the quality control section below. For species with evolutionarily significant geographic variants or subspecies, one individual from each distinct taxon was obtained. These include westslope cutthroat trout (*Oncorhynchus clarkii* ssp. *lewisi*) (McPhail 2007), Coastal and Interior lineages of bull trout (*Salvelinus confluentus*) (Taylor et al. 1999), and Morrison Creek Lamprey (*Lampetra richardsoni* var. *marifuga*) (Beamish 1987).

In most cases we obtained DNA extractions from vouchered specimens but also obtained frozen/ethanol preserved tissue from several museums. For 18 taxa with Institution = PBS in Suppl. material 1, Fisheries and Oceans Canada contractors obtained voucher specimens during other research surveys. Most species did not require a permit for collection with the exception of *Rhinichthys osculus*, collected under SARA permit 20-PPAC-00007. *Stenodus leucichthys* tissue was collected by the Teslin Tlingit Council and *Salmo salar* was collected from research aquaria at the Pacific Biological Station in Nanaimo, BC. No permit was necessary for either of these collections.

Shotgun DNA sequencing, mitogenome and rDNA assemblies, and annotation

Total genomic DNA was extracted from fin or muscle tissue using the Qiagen DNeasy Blood and Tissue kit and quantified using Quant-iT™ PicoGreen Assay (ThermoFisher). Input amounts normalized to 10 ng were used to build Illumina DNA libraries, which were sequenced on a NovaSeq SP flow cell (2 × 250 bp) at a

target sequencing depth of 5 million reads per sample. For the subset of samples with mitochondrial genome and nuclear rDNA cistron coverage below 20-fold after the first run, the same library was sequenced using another Illumina NovaSeq SP flow cell (2 × 250 bp kit) to increase read depth by 1 to 14M reads per sample. For the subset of samples for which mitochondrial genome assembly was not possible after the first run, the original genomic DNA was used in a secondary independent Illumina DNA library preparation with minor modifications for low DNA input. This was then sequenced using the Illumina NovaSeq SP flow cell (2 × 250 bp kit) at a target sequencing depth of 10 million reads per sample.

Raw sequencing data were demultiplexed and trimmed of indices using Illumina's bcl2fastq (version 2.20.0.422) software. For each sample, trimmomatic (version 0.39) (Bolger et al. 2014) was used to remove Illumina adapters and trim low-quality base calls from the ends of reads. We used several de novo assembly/annotation software toolkits since no single tool was successful at analyzing all samples. GetOrganelle (version 1.7.4.1) (Jin et al. 2020) was used to assemble paired-end reads into whole mitogenomes (or scaffolds when complete assembly was not possible); then MitoZ (version 2.3) (Meng et al. 2019) was used to annotate assemblies or scaffolds. Alternatively, some samples were assembled and annotated using MitoFlex (version 0.2.9) (Li et al. 2021), and others were assembled using the de novo assembler ABySS (version 2.2.5) (Simpson et al. 2009) and then annotated using MitoZ. No single mitogenome assembly/annotation toolkit was optimal for all the species. Therefore, we ran MitoFlex and MitoZ on all samples and chose to retain those that had the most complete assemblies (i.e., no gaps) and best annotations (i.e., most genes annotated). In several instances neither of these tools were able to assemble a complete genome. For these cases we used the tool GetOrganelle (and in one instance, Abyss) to perform the assembly, followed by MitoFlex to perform the gene annotations. Nuclear rDNA regions were annotated using the software barrnap (version 0.9) (Seeman 2009). We included all results where the length was at least 95% complete.

Quality assurance

To ensure traceability of sequence data to physical voucher specimens and minimize the likelihood of sequencing misidentified material, tissues to be sequenced were predominantly sourced from museum collections, as follows: the Royal Ontario Museum (n = 45); the University of British Columbia's Beaty Biodiversity Museum (n = 13); and the University of Washington Burke Museum Ichthyology Collection (n = 5). Exceptions to this included 9 tissue samples from the Beaty Biodiversity Museum, collected and identified by fish collection director Dr. Eric B. Taylor (E. Taylor, pers. comm.); six of which have voucher specimens that are not catalogued and four of which had no voucher specimen (both bull trout lineages, *Salvelinus confluentus*; lake trout, *Salvelinus namaycush*; and longnose sucker, *Catostomus catostomus*). Tissue from the inconnu (*Stenodus leucichthys*), collected by the Teslin Tlingit Council, also does not have a whole voucher specimen but has a tissue voucher housed at the Pacific Biological Station (Nanaimo, BC). The remainder of voucher specimens were collected by Fisheries and Oceans Canada (n = 19) and are currently being catalogued at the Royal British Columbia Museum. All voucher information is included in Suppl. material 1.

To verify concordance between morphological taxonomy and molecular taxonomy, whole COI sequences from each mitogenome were manually aligned and inspected with DNA barcodes produced by Hubert et al. (2008) in the Canadian Freshwater Fish Barcode Database (BCF and BCFB projects in BOLD, n = 190 species). We used a divergence cut-off of 0.6% to confirm matches, as Hubert et al. (2008) calculated a maximum intraspecific genetic distance of 0.6% in the ~650 bp COI barcode region for Canadian freshwater fish (excluding outliers; note that average interspecific divergence was a much higher 7.5%[57]).

The exception was lampreys (Petromyzontidae), which are not in the Canadian Freshwater Fish Barcode Database (Hubert et al. 2008), and have very little interspecific mitogenomic variation. Thus we took a multi-faceted approach to verify the genetic identity of the five species in NAMERS, including first verifying that each NAMERS sequence matched a record in GenBank derived from a voucher specimen. The *Entosphenus* genus contains *E. tridentatus* and *E. macrostomus*, the latter of which is found only in Cowichan Lake, hence we use sampling location to identify this specimen. *E. tridentatus* is widespread and the NAMERS specimen was collected in marine waters and identified by presence of three prominent teeth by lamprey expert Joy Wade (Joy Wade, pers., comm.). The *Lampetra* genus contains geographically isolated *L. richardsoni* var. *marifuga*, for which the NAMERS specimen was collected within its range at Morrison Creek, BC, Canada. It also contains *L. ayresii* and *L. richardsoni* which were differentiated based on their non-overlapping habitat use: the former was collected as an adult in marine waters whereas the latter was in freshwater.

A mitogenome portal for applied eDNA metabarcoding

NAMERS sequences and associated metadata were deposited in GenBank (under the BC Freshwater Fish Genome Project) and in a newly developed, purpose-built online mitogenome and nuclear rDNA cistron sequence data portal specifically for applied eDNA metabarcoding (https://namers.ca). Specific functionalities of the portal are summarized in Results.

Assessments of amplification and taxonomic resolution efficiencies of genetic markers are critical for sound applied eDNA metabarcoding study design as both are key determinants of success (Taberlet et al. 2012; Deagle et al. 2014). To illustrate the value of mitogenome data for marker selection, we used NAMERS to evaluate the efficacy of existing mitochondrial fish metabarcoding markers for both amplifying targets and providing species level resolution. We selected 19 eDNA metabarcoding markers targeting teleost fish from across four mitochondrial gene regions (12S, n = 7; 16S, n = 5; COI, n = 4; and cytochrome b (CYTB), n = 3) from the literature, all of which amplify targets smaller than the maximum allowable length for Illumina sequencing kits (Table 1). Each primer pair was aligned and manually checked for the number of mismatches with each species; degenerate bases in primers were scored as mismatches if none of the nucleotides coded for by the degeneracy matched the species nucleotide at that site. Amplicons for each species were generated and checked for uniqueness using two methods in R (R Core Team 2018). First, the haplotype function from the haplotypes package (Atkas 2020) identified unique haplotypes based on genetic distance calculated using dist.dna with model “raw” and pairwise deletion of indels in the ape package (Paradis and

Table 1. Information on markers assessed for primer mismatches and species level resolution in 82 freshwater fish species. [1]Markers presented at the Family level in Fig. 2.

Gene	Primer Name	Reference	Forward Sequence (5'-3')	Reverse Sequence (5'-3')	Amplicon Range (bp)	Maximum mismatches for F/R primers
12S	Teleo[1]		ACACCGCCCGTCACTCT	CTTCCGGTACACTTACCATG	61–64	7/1
	Teleo2		AAACTCGTGCCAGCCACC	GGGTATCTAATCCCAGTTTG	164–177	1/1
	MiFishU[1]		GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG	168–181	3/2
	AcMDB07[1]	(Bylemans et al. 2018)	GCCTATATACCGCCGTCG	GTACACTTACCATGTTACGACTT	241–282	1/1
	Am12S	(Evans et al. 2016)	AGCCACCGCGGTTATACG	CAAGTCCTTTGGGTTTTAAGC	237–253	1/3
	Ac12S	(Evans et al. 2016)	ACTGGGATTAGATACCCCACTATG	GAGAGTGACGGGCGGTGT	370–392	2/1
	12S_V5	(Riaz et al. 2011)	ACTGGGATTAGATACCCC	TAGAACAGGCTCCTCTAG	89–107	1/1
	Ac16S	(Evans et al. 2016)	CCTTTTGCATCATGATTTAGC	CAGGTGGCTGCTTTTAGGC	321–341	2/5
	Shaw16S	(Shaw et al. 2016)	CGAGAAGACCCTWTGGAGCTTIAG	GGTCGCCCCAACCRAAG	56–80	3/3
	Vert 16S[1]	(Vences et al. 2016)	AGACGAGAAGACCCYTGGAGCTT	GATCCAACATCGAGGTCGTAA	237–278	1/0
	L2513/H2714[1]	(Kitano et al. 2007)	GCCTGTTTACCAAAAACATCAC	CTCCATAGGGTCTTCTCGTCTT	201–205	2/1
	Fish16SF-16S2R	(Berry et al. 2017)	GACCCTATGGAGCTTTAGAC	CGCTGTTATCCCTADRGTAACT	188–216	5/1
CO1	SeaDNA-short[1]	(Collins et al. 2019)	GGAGGCTTTGGMAAYTGRYT	GGGGGAAGAARYCARAARCT	55	4/4
	LerayXT[1]	(Wangensteen et al. 2018)	GGWACWRGWTGRACWITITAYCCYCC*	TAIACYTCIGGRTGICCRAARAAYCA*	313	4/0
	seaDNA-mid	(Collins et al. 2019)	GGAGGCTTTGGMAAYTGRYT	TAGAGGRGGGTARACWGTYCA	130	4/5
	Minibar	(Meusnier et al. 2008)	TCCACTAATCACAARGATATTGGTAC	GAAAATCATAATGAAGGCATGAGC	127	5/8
CYTB	Minamoto-fish[1]	(Minamoto et al. 2012)	TTCCTAGCCATACAYTAYAC	GGTGGCKCCTCAGAAGGACATTTGKCCYCA	235	4/8
	FishCBL/ FishCBR	(Thomsen et al. 2012)	TCCTTTTGAGGCGCTACAGT	GGAATGCGAAGAATCGTGTT	90	9/6
	Fish2CBL/ Fish2bCBR	(Thomsen et al. 2012)	ACAACTTCACCCCTGCAAAC	GATGGCGTAGGCAAACAAGA	N/A	6/6

Schliep 2018). This method did not identify unique haplotypes due to indels, therefore, the GetHaplo function in the SIDIER package (Muñoz-Pajares 2021) was used to identify these. Each unique haplotype was manually checked and considered as providing species level resolution if it was at least 1 bp different (including indels) from another species.

Results

Shotgun DNA sequencing, assemblies, and annotation

Mitogenomic data were generated here for an estimated 92.3% (85/92) of all freshwater fish taxa present in our target geographic area of BC, Canada, representing 49 genera and 19 families. Sequencing success rates were high; 82 of 85 taxa sequenced returned complete or near complete (missing one gene or few partial genes) mitogenomes and a further three returned partial mitogenomes. Thus the final data set is comprised of complete or near complete mitogenome sequences for ~89% of all freshwater fish taxa in BC (82/92) and partial mitogenomes for an additional two species and one lineage. Mitogenome

sequencing depth ranged from 1.4 to 2249.9 (median 101.2) and mitogenome length ranged from 14.198–18.141 kbp (median 16.634 kbp). All species for which the full mitogenome was constructed contained 13 protein-coding genes (COX1 – COX3, CYTB, ND1 – ND6, ND4L, ATP6, and ATP8), 22 tRNA genes, and two rRNA genes (small and large rRNA subunits). Full nrDNA cistrons containing 5.8S, 18S, and 28S regions were sequenced for 70 species, with sequencing depth ranging from 16.5 to 1340.8 with a median of 413.4. Full details on mitogenome and nrDNA data are in Suppl. materials 1, 3.

Genetic verification of morphological taxonomy

The morphological taxonomy of each specimen in NAMERS was verified using genetic identification by the COI barcode region in almost all instances, with a few exceptions as follows. The candidate Umatilla dace (Cyprinidae; *Rhinichthys umatilla*) specimen was a misidentified sucker (Family Catostomidae), which is highly plausible given the difficulty identifying juvenile fish, and hence excluded. For lampreys, COI and 12S genes were invariable within genera; however, the 16S gene differentiated *Entosphenus* species and the cytochrome b (CYTB) gene differentiated *Lampetra* species (excluding the Morrison creek variant), by a single base in all cases. The ND4 gene had highest genetic variation among lamprey species, with three base changes between the two *Entosphenus* species and two changes between the two *Lampetra* species (again excluding the Morrison Creek variant); suggesting that ND4 may be a candidate gene for species specific markers in this family.

NAMERS portal specifications

Whole mitogenomes, annotated mitochondrial genes, and annotated nuclear ribosomal genes are available to view and download in FASTA format on the new NAMERS portal. The main database page offers a table of 86 species grouped by increasing taxonomic levels. Users can highlight any taxonomic level to view available sequence data for all included taxa, from individual species to family, and can easily customize batches of particular genes or taxa for downloading in FASTA format. They can also highlight particular genes of interest or the complete mitogenome for automatic alignments (using MUSCLE, Edgar 2004) of selected genes and taxa of interest; the alignment will automatically update when the user engages or disengages genes or taxa. Clicking on an individual species will take users to a species page with mtDNA and nrDNA *de novo* assembly statistics, mitogenome circular sketch map, and link to download the full mitogenome.

The number of primer mismatches and the proportion of species resolved for 19 published fish and vertebrate metabarcoding markers was assessed using the NAMERS database (Table 1 and Fig. 1). Due to low genetic variation within Petromyzontidae, no metabarcoding marker could differentiate species (Fig. 2), therefore we report species resolution rates with and without lamprey included (Fig. 1). Across all 86 taxa (representing 19 families), the number of mismatches in forward and reverse primers was highly variable, ranging from 0 to 8, with 12S and 16S primers generally having fewer than COI and CYTB (Fig. 1). The single marker with the fewest overall mismatches was Vert16S (Vences et al. 2016). Excluding lampreys, species resolution rates among markers tested varied from 75% to 100% (Fig. 1).

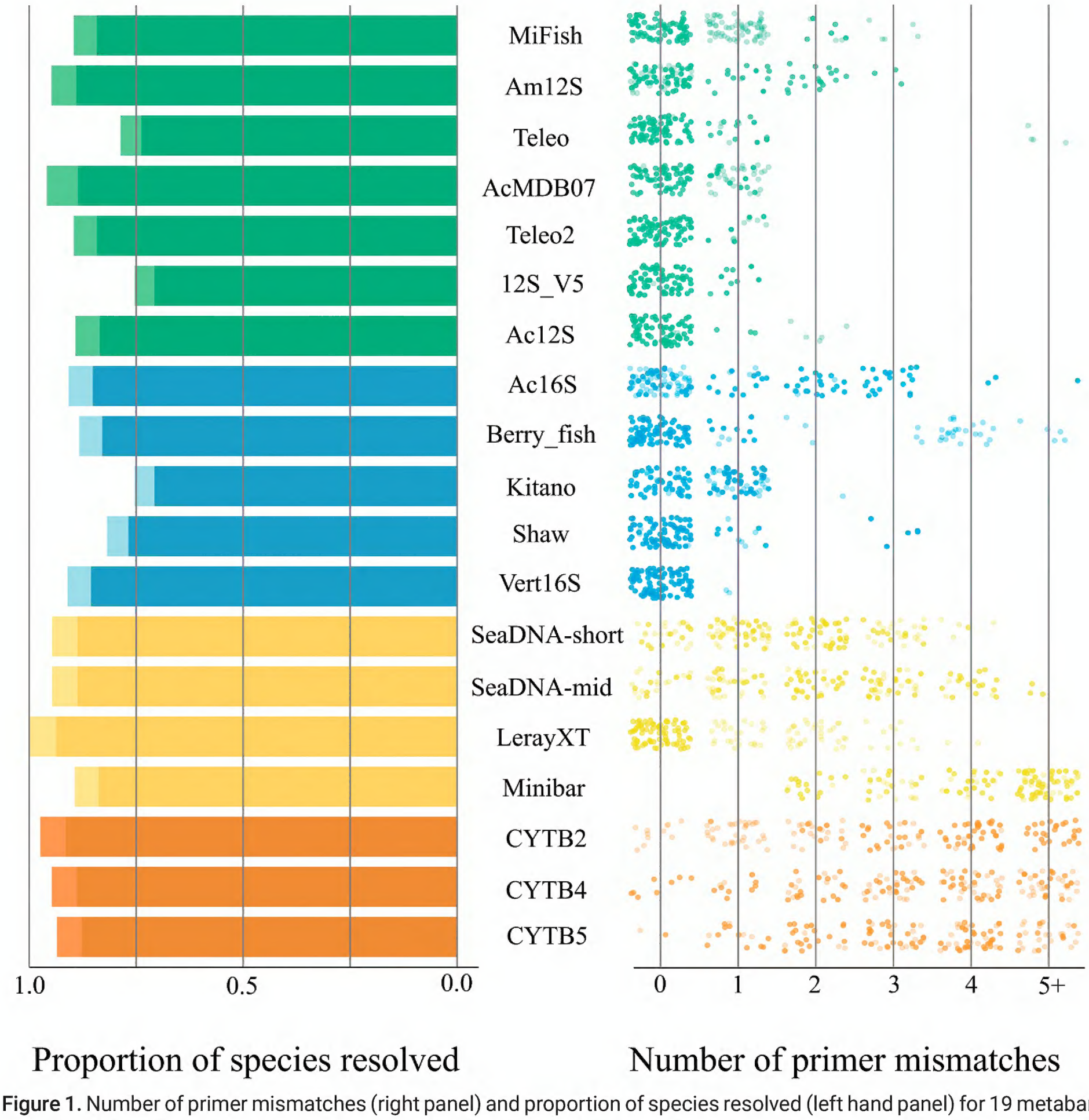

Figure 1. Number of primer mismatches (right panel) and proportion of species resolved (left hand panel) for 19 metabarcoding markers from four gene regions, generated using all species in NAMERS (n = 86). Forward primer mismatches are depicted by dark circles and reverse primer mismatches by light circles. Superscripts indicate the number of ambiguous bases in the forward and reverse primers, respectively. Markers with no superscripts have no ambiguous bases in either primer. Darker bar area shows species resolution when lamprey are included and the entire bar when they are excluded.

Achieving high confidence species level resolution within a family will at times be more important than surveying across all families, as some applied eDNA metabarcoding surveys will focus on lower taxonomic groups only, depending on the specific survey aim. The number of primer mismatches and proportion of species resolved for families (n = 6) with more than five species (up to 63 species) are shown in Fig. 2. Species resolution rates within families ranged markedly among markers tested, from zero in lampreys (Petromyzontidae, n = 5) to 100% in sunfishes (Centrarchidae, n = 5; Fig. 2). Noteworthy is that in four of the six families in NAMERS with at least five species, the Vert16S marker resolved 100% of species with zero primer mismatches (Fig. 2).

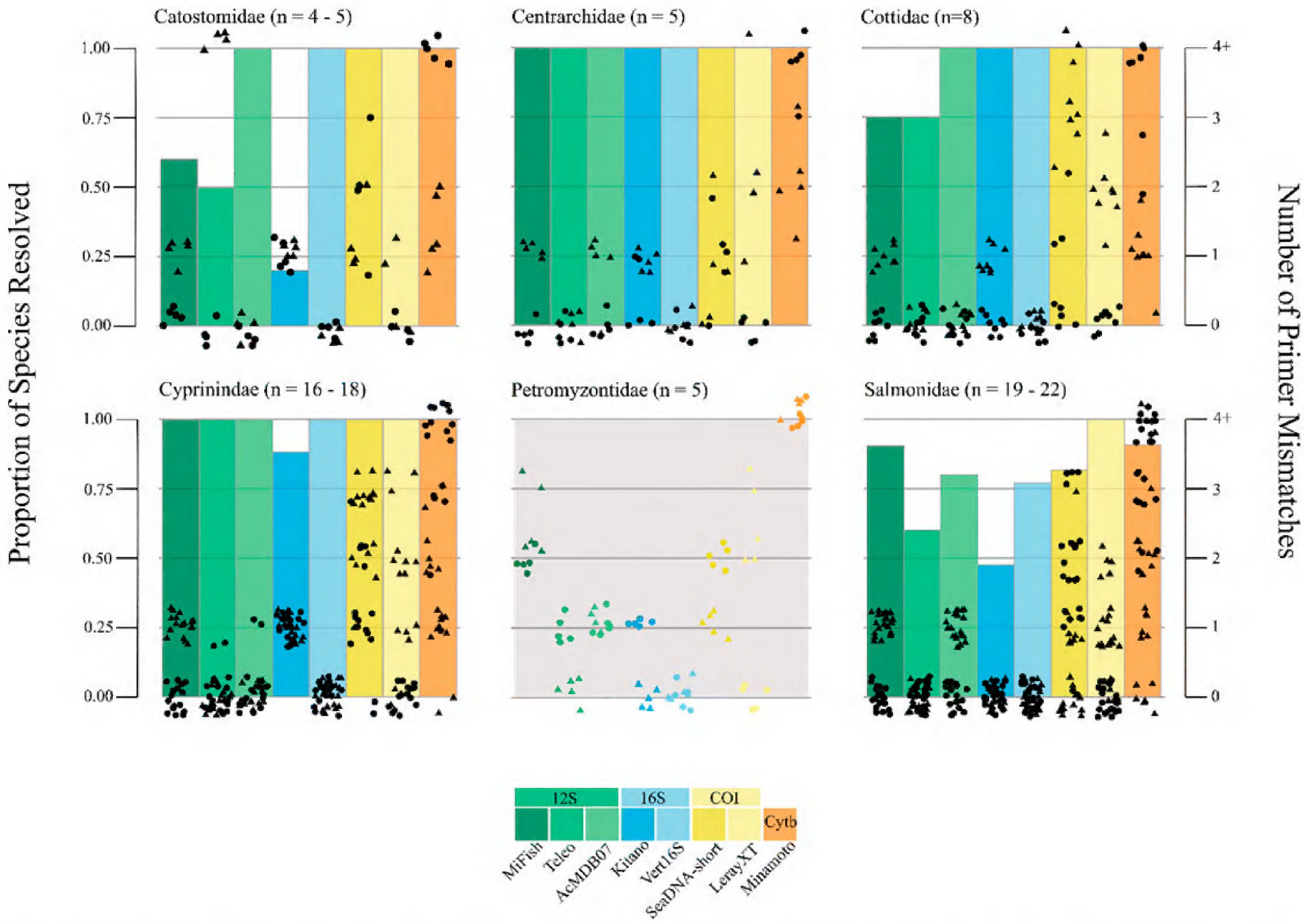

Figure 2. Family level plots of the number of primer mismatches, depicted by solid black symbols and the right-hand y-axis (forward primer = black triangle, reverse primer = black circle), and the proportion of species resolved by unique amplicons defined as a minimum of 1 bp difference including indels, depicted by coloured bars and the left hand y-axis. Plots are for all families in the NAMERS database with a minimum of five species and a subset of eight of the primers tested in Fig. 1. No species in Petromyzontidae were resolved by any marker. The range in number of species per family used in the analysis reflects missing data; some species had some missing or partial genes available for analysis (See Suppl. materials 1, 2).

Discussion

Current unprecedented rates of global change and biodiversity loss demand innovative and efficient tools for monitoring and managing ecosystems (IBPES 2019). eDNA metabarcoding represents a powerful tool in this regard, offering enhanced detection capabilities, efficiency, and a broad range of applications that can improve our understanding and management of a rapidly changing biosphere (Sahu et al. 2023). By leveraging this technology, we can better address the challenges posed by global environmental changes and work towards preserving biodiversity. However, minimum quality criteria and communicating sources of uncertainty are vital for translating eDNA based monitoring results into policy and sound decision-making (Darling and Mahon 2011; Mauvisseau et al. 2019; Mathieu et al. 2020; Darling et al. 2021).

As defined earlier, the term applied eDNA metabarcoding encompasses unique quality- and confidence-related needs that come with translating this eDNA method into practical application. Here we introduced a framework combining three foundational concepts of quality, completeness, and accessibility, to improve reference sequence repositories to meet the unique needs of applied

eDNA metabarcoding. Although NAMERS is a region-specific database exemplifying this framework, these concepts can be applied to new iterations of reference sequence databases around the world as this technology is increasingly integrated into management models. We advocate for establishing large-scale databases with long-term funding models that incorporate the foundational concepts described here. The regionally-focused NAMERS database may not include the taxonomic breadth for studies of anthropogenic-mediated introductions or climate-mediated shifts in species distributions. However, there are clear advantages of the three foundational pillars for reference DNA data demonstrated in NAMERS that are not present in any other single existing repository.

Completeness

Species richness can be underestimated by indiscriminate application of metabarcoding markers without a full understanding of their specificity for target groups or level of species coverage within the reference library used for taxonomy assignments (Gold et al. 2021). Primer specificity is vital to understanding the likelihood of false negatives as the more mismatches between a primer and its target sequence will decrease the ability for the primer to bind, and this effect is more pronounced in the 3' region of the primer. By having complete mitogenomes available as one of the pillars of NAMERS (completeness), primer mismatch can be carefully assessed against target taxa to determine the probability of false negatives. This data driven survey design enables informed decision making at the outset and increases confidence in non-detection as a true signal rather than due to low primer specificity. Overall patterns in primer mismatches were revealed by the mismatch analyses here. On average, primers in protein coding COI and CYTB genes had more mismatches (e.g. decreased specificity) compared to primers in the 12S and 16S ribosomal DNA genes. The decreased specificity of COI led to low levels of reproducibility due to non-specific amplification compared to 12S in a metabarcoding study of marine and freshwater fishes in the UK (Collins et al. 2019).

Taxonomically complete reference libraries like NAMERS also allow species resolution to be assessed as part of the survey design phase, which may be especially important when specific taxonomic groups are targeted. In our analyses, even though average species resolution was greater for the protein coding COI and CYTB genes as expected, the 12S and 16S genes had lower average rates of primer mismatch and would therefore likely recover more freshwater fish species when conducting taxonomically broad surveys. Family- or genus-level species resolution is likely to be a more common priority at multiple levels of government, even if these types of studies are perhaps less represented in the literature. These specific eDNA metabarcoding applications are often in the conservation and invasive species areas (Feist and Lance 2021). Comparing the analyses performed here, species resolution rates could inform marker choice differently than primer specificity. For example, Salmonids are a culturally and economically important species in BC with a high level of conservation priority, and only one marker we assessed could differentiate all species. Although the LerayXT marker had up to four mismatches across all families (Fig. 1 and Table 1), there are only a maximum of two mismatches within the Salmonids (Fig. 2), making it a viable marker for studies focusing on this family but not a viable marker for more general freshwater fish studies.

Quality

Reference DNA sequences are a challenging element of the eDNA metabarcoding workflow for which to satisfy quality control and assurance criteria for sufficient confidence in results (Taberlet et al. 2012; Schroeter et al. 2020; Zaiko et al. 2022), transparency of methods, and transparency of data (Zaiko et al. 2018). Global genetic repositories do not have the level of curation necessary for defensible species assignments from environmental sequences (Zaiko et al. 2022), and are problematic with respect to low quality control, reproducibility, transferability, and ease of regular updating (Collins et al. 2021). Whilst some level of error correction is achievable (Collins et al. 2021), several elements that lead to errors cannot be corrected, including uneven taxonomic coverage, undocumented specimen metadata, and lack of voucher availability or quality control processes. Several pipelines are available for downloading and curating GenBank data to create eDNA metabarcoding reference libraries (e.g. Collins et al. 2021; Claver et al. 2023). Removing erroneous sequences is the most challenging aspect of curation as phylogenetic methods are cumbersome with large data sets and short amplicons contain fewer informative sites. Sequence similarity is generally used to determine erroneous sequences; however, this metric can be problematic for widely distributed species with high levels of intraspecific variation. Without structured geographic metadata required for sequence records, accurate and complete removal of erroneous sequences is not achievable.

Further, since eDNA metabarcoding tools will often be multi-marker and implemented at local and regional scales, rarely global ones, the use of both single gene repositories and massive sequence databases is impractical as generating custom libraries from these is laborious and requires specialized expertise. As an example, GenBank only has patchy availability of geo-referenced vouchers because this is not a requirement for submission. Thus ensuring the traceability for specimen identity as established in NAMERS is not easily achievable in GenBank.

Accessibility

User friendly platforms and bioinformatic pipelines have been generally missing from eDNA metabarcoding research and development, yet with the potential global reach for this technology, these elements are going to become increasingly valuable for specialists and non-specialists alike. The web platform developed for NAMERS showcases several functionalities and accessibility features that set it above other leading reference DNA databases. We acknowledge some elements of the existing layout may not be scalable for larger databases but suit the regional focus well. The data table provides an overview of the species in the database and the availability of genetic data. The alignment viewer is the most advanced part of the platform, where users can choose multiple custom taxa and genes (one gene at a time), view the multiple sequence alignments, and download their custom reference DNA data sets. Other leading databases, such as MIDORI2 (Leray et al. 2022) and MitoFish (Sato et al. 2018) do not offer the same combined features. MIDORI2 lacks the ability to easily define regionally relevant taxonomic groups and MitoFish lacks the ability to define both taxonomic groups and genes of interest. Furthermore, MIDORI2 contains

the range of fragment sizes contained in GenBank and therefore is not applicable for front-end data driven design software such as UNIKSEQ (Allison et al. 2023), which relies on mitogenomes to identify novel markers. Furthermore, the NAMERS database is versioned to control for changes in taxonomic classification and additional taxa or geographic representatives. This increases transparency in data provenance for end users with respect to any changes to the database and their impact on the interpretation of metabarcoding results.

Conclusion

It is no longer far-fetched to make whole mitogenomes the new standard for reference DNA sequences given genome skimming capabilities (Hoban et al. 2022) and the steady decrease in cost-per-base for ultra-high throughput sequencing. The NAMERS repository is presented here as a framework or foundation to inspire this investment and motivate the considerable coordination that will be necessary to extend it to include larger geographic and taxonomic scales. Incorporating the elements of NAMERS into larger initiatives such as the Earth Biogenome Project (Lewin et al. 2018) would add value for eDNA practitioners. Particular elements could be searchable databases with the ability to download customized files. The NAMERS database echoes Dziedzic et al. (2023) in demonstrating that a small research group can generate a comprehensive set of high quality mitogenome reference sequences for a geographically relevant area, which can be impactful on its own or within a larger framework. We acknowledge that long-term support and continued augmentation of any database like NAMERS requires dedicated resources, and yet argue the importance of this given the increased need for metabarcoding in real-world applications. While there are inherent limitations to any survey approach, especially for large-scale and taxonomically broad biota, eDNA metabarcoding has proven "good enough" for a large number of use cases (Hajibabaei 2022; Kelly et al. 2024). This is especially important as metabarcoding coupled with deep sequencing and specific primer sets has shown to be more sensitive than single species qPCR detections, providing a superior tool for biodiversity analysis of rare targets (Westfall et al. 2022; McCarthy et al. 2023). We don't have to wait until the technology and reference DNA sequence databases are perfect to start applying it, time is of the essence and we need to generate this data at scale right now.

Acknowledgements

We thank the following people for contributions related to specimen collection and curation: Liane Stenhouse (DFO), Paul Grant (DFO), Nellie Gagné (DFO), Louise-Marie Roux (DFO), Mélanie Roy (DFO), Joy Wade (Fundy Aquaculture Services), Rick Taylor (UBC Beaty Biodiversity Museum), Jordan Rosenfeld (BC Ministry of Environment and Climate Change Strategy), Bob Hanner (University of Guelph), Daniel Heath (University of Windsor), Gavin Hanke (Royal BC Museum), Teslin Tlingit Council, Pascale Savage (Yukon Government), Caren Helbing (University of Victoria), Amelia Louden (Burke Museum), Louis Lopez (University of Victoria), Hoda Rajabi (eDNAtec), Emily Porter (eDNAtec), and Avery McCarthy (eDNAtec). The authors would also like to thank two anonymous reviewers for their valuable input.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Funding

This research was funded by Genome BC, Project #SIP26-06.

Author contributions

CLA, KMW, and SRG conceived of the study and obtained project funding. KMW prepared samples for sequencing. MH, NF, and GACS managed sequencing and performed bioinformatics. MK and GACS built the website with input from all authors. KMW and CLA wrote the manuscript with input from all authors.

Author ORCIDs

Kristen M. Westfall https://orcid.org/0000-0001-7524-7145

Data availability

All processed genetic data is available from https://namers.ca and is available in GenBank (Accession Numbers in Suppl. materials 2, 3).

References

Allison MJ, Warren RL, Lopez ML, Acharya-Patel N, Imbery JJ, Coombe L, Yang CL, Birol I, Helbing CC (2023) Enabling robust environmental DNA assay design with “unikseq” for the identification of taxon-specific regions within whole mitochondrial genomes. Environmental DNA 5(5): 1032–1047. https://doi.org/10.1002/edn3.438

Atkas C (2020) haplotypes: Manipulating DNA Sequences and Estimating Unambiguous Haplotype Network with Statistical Parsimony. https://cran.r-project.org/web/packages/haplotypes/index.html

Baird DJ, Hajibabaei M (2012) Biomonitoring 2.0: A new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. Molecular Ecology 21(8): 2039–2044. https://doi.org/10.1111/j.1365-294X.2012.05519.x

Beamish RJ (1987) Evidence that parasitic and non-parasitic life history types are produced by one population of lamprey. Canadian Journal of Fisheries and Aquatic Sciences 44(10): 1779–1782. https://doi.org/10.1139/f87-219

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, Wheeler DL (2000) GenBank. Nucleic Acids Research 28(1): 15–18. https://doi.org/10.1093/nar/28.1.15

Berry TE, Osterrieder SK, Murray DC, Coghlan ML, Richardson AJ, Grealy AK, Stat M, Bejder L, Bunce M (2017) DNA metabarcoding for diet analysis and biodiversity: A case study using the endangered Australian sea lion (*Neophoca cinerea*). Ecology and Evolution 7(14): 5435–5453. https://doi.org/10.1002/ece3.3123

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England) 30(15): 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bylemans J, Gleeson DM, Hardy CM, Furlan E (2018) Toward an ecoregion scale evaluation of eDNA metabarcoding primers: A case study for the freshwater fish biodiversity of the Murray–Darling Basin (Australia). Ecology and Evolution 8(17): 8697–8712. https://doi.org/10.1002/ece3.4387

Claver C, Canals O, de Amézaga LG, Mendibil I, Rodriguez-Ezpeleta N (2023) An automated workflow to assess completeness and curate GenBank for environmental DNA metabarcoding: The marine fish assemblage as case study. Environmental DNA 5(4): 634–647. https://doi.org/10.1002/edn3.433

Collins RA, Bakker J, Wangensteen OS, Soto AZ, Corrigan L, Sims DW, Genner MJ, Mariani S (2019) Non-specific amplification compromises environmental DNA metabarcoding with COI. Methods in Ecology and Evolution 10(11): 1985–2001. https://doi.org/10.1111/2041-210X.13276

Collins RA, Trauzzi G, Maltby KM, Gibson TI, Ratcliffe FC, Hallam J, Rainbird S, Maclaine J, Henderson PA, Sims DW, Mariani S, Genner MJ (2021) Meta-Fish-Lib: A generalised, dynamic DNA reference library pipeline for metabarcoding of fishes. Journal of Fish Biology 99(4): 1446–1454. https://doi.org/10.1111/jfb.14852

Cordier T, Alonso-Sáez L, Apothéloz-Perret-Gentil L, Aylagas E, Bohan DA, Bouchez A, Chariton A, Creer S, Frühe L, Keck F, Keeley N, Laroche O, Leese F, Pochon X, Stoeck T, Pawlowski J, Lanzén A (2021) Ecosystems monitoring powered by environmental genomics: A review of current strategies with an implementation roadmap. Molecular Ecology 30(13): 2937–2958. https://doi.org/10.1111/mec.15472

Creer S, Deiner K, Frey S, Porazinska D, Taberlet P, Thomas WK, Potter C, Bik HM (2016) The ecologist's field guide to sequence-based identification of biodiversity. Methods in Ecology and Evolution 7(9): 1008–1018. https://doi.org/10.1111/2041-210X.12574

Cristescu ME, Hebert PDN (2018) Uses and misuses of environmental DNA in biodiversity science and conservation. Annual Review of Ecology, Evolution, and Systematics 49(1): 209–230. https://doi.org/10.1146/annurev-ecolsys-110617-062306

Darling JA, Mahon AR (2011) From molecules to management: Adopting DNA-based methods for monitoring biological invasions in aquatic environments. Environmental Research 111(7): 978–988. https://doi.org/10.1016/j.envres.2011.02.001

Darling JA, Pochon X, Abbott CL, Inglis GJ, Zaiko A (2020) The risks of using molecular biodiversity data for incidental detection of species of concern. Diversity & Distributions 26(9): 1116–1121. https://doi.org/10.1111/ddi.13108

Darling JA, Jerde CL, Sepulveda AJ (2021) What do you mean by false positive? Environmental DNA 3(5): 879–883. https://doi.org/10.1002/edn3.194

Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P (2014) DNA metabarcoding and the cytochrome c oxidase subunit I marker: Not a perfect match. Biology Letters 10(9): 20140562. https://doi.org/10.1098/rsbl.2014.0562

Dziedzic E, Sidlauskas B, Cronn R, Anthony J, Cornwell T, Friesen TA, Konstantinidis P, Penaluna BE, Stein S, Levi T (2023) Creating, curating and evaluating a mitogenomic reference database to improve regional species identification using environmental DNA. Molecular Ecology Resources 23(8): 1880–1904. https://doi.org/10.1111/1755-0998.13855

Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32(5): 1792–1797. https://doi.org/10.1093/nar/gkh340

European Parliament & Council of the European Union (2008) Directive establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive). Directive 2008/56/EC. European Commission. Joint

Research Centre. Institute for Environment and Sustainability. Marine Strategy Framework Directive - Competence Centre. https://mcc.jrc.ec.europa.eu/main/index.py

Evans NT, Olds BP, Renshaw MA, Turner CR, Li Y, Jerde CL, Mahon AR, Pfrender ME, Lamberti GA, Lodge DM (2016) Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. Molecular Ecology Resources 16(1): 29–41. https://doi.org/10.1111/1755-0998.12433

Feist SM, Lance RF (2021) Advanced molecular-based surveillance of quagga and zebra mussels: A review of environmental DNA/RNA (eDNA/eRNA) studies and considerations for future directions. NeoBiota 66: 117–159. https://doi.org/10.3897/neobiota.66.60751

Ficetola GF, Boyer F, Valentini A, Bonin A, Meyer A, Dejean T, Gaboriaud C, Usseglio-Polatera P, Taberlet P (2021) Comparison of markers for the monitoring of freshwater benthic biodiversity through DNA metabarcoding. Molecular Ecology 30(13): 3189–3202. https://doi.org/10.1111/mec.15632

Gold Z, Curd EE, Goodwin KD, Choi ES, Frable BW, Thompson AR, Walker Jr HJ, Burton RS, Kacev D, Martz LD, Barber PH (2021) Improving metabarcoding taxonomic assignment: A case study of fishes in a large marine ecosystem. Molecular Ecology Resources 21(7): 2546–2564. https://doi.org/10.1111/1755-0998.13450

Goodwin KD, Aiello CM, Weise M, Edmondson M, Fillingham K, Allen D, Amerson A, Barton ML, Benson A, Canonico G, Gold Z, Gumm J, Hunter M, Joffe N, Lance R, Larkin A, Letelier R, Lipsky C, McCoskey D, Morrison C, Clark K, Darling JA, Demery A-J, Everett M, Fletcher-Hoppe C, Nichols KM, Parsons KM, Price J, Puglise K, Scholl K, Schwartz MK, Sepulveda A, Shannon J, Turner W, White T (2024) National Aquatic Environmental DNA Strategy. US Geological Survey Report, Wetland and Aquatic Research Centre, 70255545. https://www.usgs.gov/publications/national-aquatic-environmental-dna-strategy

Hajibabaei M (2022) Demystifying eDNA validation. Trends in Ecology & Evolution 37(10): 826–828. https://doi.org/10.1016/j.tree.2022.06.015

Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ (2011) Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. PLOS ONE 6(4): e17497. https://doi.org/10.1371/journal.pone.0017497

Hering D, Borja A, Jones JI, Pont D, Boets P, Bouchez A, Bruce K, Drakare S, Hanfling B, Kahlert M, Leese F, Meissner K, Mergen P, Reyjol Y, Segurado P, Vogler A, Kelly M (2018) Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. Water Research 138: 192–205. https://doi.org/10.1016/j.watres.2018.03.003

Hoban ML, Whitney J, Collins AG, Meyer C, Murphy KR, Reft AJ, Bemis KE (2022) Skimming for barcodes: Rapid production of mitochondrial genome and nuclear ribosomal repeat reference markers through shallow shotgun sequencing. PeerJ 10: e13790. https://doi.org/10.7717/peerj.13790

Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E, Burridge M, Watkinson D, Dumont P, Curry A, Bentzen P, Zhang J, April J, Bernatchez L (2008) Identifying Canadian freshwater fishes through DNA barcodes. PLOS ONE 3(6): e2490. https://doi.org/10.1371/journal.pone.0002490

IBPES (2019) Summary for Policymakers of the Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.

Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ (2020) GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biology 21(1): 241. https://doi.org/10.1186/s13059-020-02154-5

Kelly RP, Lodge DM, Lee KN, Theroux S, Sepulveda AJ, Scholin CA, Craine JM, Andruszkiewicz Allan E, Nichols KM, Parsons KM, Goodwin KD, Gold Z, Chavez FP, Noble RT, Abbott CL, Baerwald MR, Naaum AM, Thielen PM, Simons AL, Jerde CL, Duda JJ, Hunter ME, Hagan JA, Meyer RS, Steele JA, Stoeckle MY, Bik HM, Meyer CP, Stein E, James KE, Thomas AC, Demir-Hilton E, Timmers MA, Griffith JF, Weise MJ, Weisberg SB (2024) Toward a national eDNA strategy for the United States. Environmental DNA 6: e432. https://doi.org/10.1002/edn3.432

Kitano T, Umetsu K, Tian W, Osawa M (2007) Two universal primer sets for species identification among vertebrates. International Journal of Legal Medicine 121(5): 423–427. https://doi.org/10.1007/s00414-006-0113-y

Leese F, Bouchez A, Abarenkov K, Altermatt F, Borja A, Bruce K, Ekrem T, Čiampor F, Čiamporová-Zaťovičová Z, Costa FO, Duarte S, Elbrecht V, Fontaneto D, Franc A, Geiger MF, Hering D, Kahlert M, Stroil BK, Kelly M, Keskin E, Liska I, Mergen P, Meissner K, Pawlowski J, Penev L, Reyjol Y, Rotter A, Steinke D, van der Wal B, Vitecek S, Zimmermann J, Weigand AM (2018) Chapter Two - Why We Need Sustainable Networks Bridging Countries, Disciplines, Cultures and Generations for Aquatic Biomonitoring 2.0: A Perspective Derived From the DNAqua-Net COST Action. In: Bohan DA, Dumbrell AJ, Woodward G, Jackson M (Eds) Advances in Ecological Research 58, 63–99. https://doi.org/10.1016/bs.aecr.2018.01.001

Leray M, Knowlton N, Machida RJ (2022) MIDORI2: A collection of quality controlled, preformatted, and regularly updated reference databases for taxonomic assignment of eukaryotic mitochondrial sequences. Environmental DNA 4(4): 894–907. https://doi.org/10.1002/edn3.303

Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, Durbin R, Edwards SV, Forest F, Gilbert MTP, Goldstein MM, Grigoriev IJ, Hackett KJ, Haussler D, Jarvis ED, Johnson WE, Patrinos A, Richards S, Castilla-Rubio JC, van Sluys M-A, Soltis PS, Xu X, Yang H, Zhang G (2018) Earth BioGenome Project: Sequencing life for the future of life. Proceedings of the National Academy of Sciences of the United States of America 115(17): 4325–4333. https://doi.org/10.1073/pnas.1720115115

Li JY, Li WX, Wang AT, Zhang Y (2021) MitoFlex: An efficient, high-performance toolkit for animal mitogenome assembly, annotation and visualization. Bioinformatics (Oxford, England) 37(18): 3001–3003. https://doi.org/10.1093/bioinformatics/btab111

Mathieu C, Hermans SM, Lear G, Buckley TR, Lee KC, Buckley HL (2020) A Systematic Review of Sources of Variability and Uncertainty in eDNA Data for Environmental Monitoring. Frontiers in Ecology and Evolution 8: 135. https://doi.org/10.3389/fevo.2020.00135

Mauvisseau Q, Burian A, Gibson C, Brys R, Ramsey A, Sweet M (2019) Influence of accuracy, repeatability and detection probability in the reliability of species-specific eDNA based approaches. Scientific Reports 9(1): 580. https://doi.org/10.1038/s41598-018-37001-y

McCarthy A, Rajabi H, McClenaghan B, Fahner NA, Porter E, Singer GAC, Hajibabaei M (2023) Comparative analysis of fish environmental DNA reveals higher sensitivity achieved through targeted sequence-based metabarcoding. Molecular Ecology Resources 23(3): 581–591. https://doi.org/10.1111/1755-0998.13732

McElroy ME, Dressler TL, Titcomb GC, Wilson EA, Deiner K, Dudley TL, Eliason EJ, Evans NT, Gaines SD, Lafferty KD, Lamberti GA, Li Y, Lodge DM, Love MS, Mahon AR, Prefnder ME, Renshaw MA, Selkoe KA, Jerde CL (2020) Calibrating Environmental DNA Metabarcoding to Conventional Surveys for Measuring Fish Species Richness. Frontiers in Ecology and Evolution 8: 276. https://doi.org/10.3389/fevo.2020.00276

McPhail JD (2007) The Freshwater Fishes of British Columbia. University of Alberta, Calgary Alberta, 620 pp. https://doi.org/10.1515/9780888648532

Meng G, Li Y, Yang C, Liu S (2019) MitoZ: A toolkit for animal mitochondrial genome assembly, annotation and visualization. Nucleic Acids Research 47(11): e63. https://doi.org/10.1093/nar/gkz173

Meusnier I, Singer GAC, Landry JF, Hickey DA, Hebert PDN, Hajibabaei M (2008) A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9(1): 214. https://doi.org/10.1186/1471-2164-9-214 [online]

Minamoto T, Yamanaka H, Takahara T, Honjo MN, Kawabata Z (2012) Surveillance of fish species composition using environmental DNA. Limnology 13(2): 193–197. https://doi.org/10.1007/s10201-011-0362-4

Muñoz-Pajares AJ (2021) Nosidier: Substitution and Indel Distances to Infer Evolutionary Relationships. https://rdrr.io/cran/sidier/

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover B, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD (2016) Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Research 44(D1): D733–D745. https://doi.org/10.1093/nar/gkv1189

Paradis E, Schliep K (2018) ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics (Oxford, England) 35(3): 526–528. https://doi.org/10.1093/bioinformatics/bty633

Piñol J, Mir G, Gomez-Polo P, Agustí N (2015) Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. Molecular Ecology Resources 15(4): 819–830. https://doi.org/10.1111/1755-0998.12355

R Core Team (2018) R: A Language and Environment for Statistical Computing.

Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System: Barcoding. Molecular Ecology Notes 7(3): 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, Coissac E (2011) ecoPrimers: Inference of new DNA barcode markers from whole genome sequence analysis. Nucleic Acids Research 39(21): e145. https://doi.org/10.1093/nar/gkr732

Ruppert KM, Kline RJ, Rahman MS (2019) Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. Global Ecology and Conservation 17: e00547. https://doi.org/10.1016/j.gecco.2019.e00547

Ruskey JA, Taylor EB (2016) Morphological and genetic analysis of sympatric dace within the *Rhinichthys cataractae* species complex: A case of isolation lost. Biological Journal of the Linnean Society. Linnean Society of London 117(3): 547–563. https://doi.org/10.1111/bij.12657

Sahu A, Kumar N, Singh CP, Singh M (2023) Environmental DNA (eDNA): Powerful technique for biodiversity conservation. Journal for Nature Conservation 71: 126325. https://doi.org/10.1016/j.jnc.2022.126325

Sato Y, Miya M, Fukunaga T, Sado T, Iwasaki W (2018) MitoFish and MiFish Pipeline: A mitochondrial genome database of fish with an analysis pipeline for environmental DNA metabarcoding. Molecular Biology and Evolution 35(6): 1553–1555. https://doi.org/10.1093/molbev/msy074

Schenekar T (2022) The current state of eDNA research in freshwater ecosystems: Are we shifting from the developmental phase to standard application in biomonitoring? Hydrobiologia 850: 1263–1282. https://doi.org/10.1007/s10750-022-04891-z

Schroeter JC, Maloy AP, Rees CB, Bartron ML (2020) Fish mitochondrial genome sequencing: Expanding genetic resources to support species detection and biodiversity monitoring using environmental DNA. Conservation Genetics Resources 12(3): 433–446. https://doi.org/10.1007/s12686-019-01111-0

Seeman T (2009) barnnap. https://github.com/tseemann/barrnap

Shaw JLA, Clarke LJ, Wedderburn SD, Barnes TC, Weyrich LS, Cooper A (2016) Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system. Biological Conservation 197: 131–138. https://doi.org/10.1016/j.biocon.2016.03.010

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol İ (2009) ABySS: A parallel assembler for short read sequence data. Genome Research 19(6): 1117–1123. https://doi.org/10.1101/gr.089532.108

Stein ED, Jerde CL, Allan EA, Sepulveda AJ, Abbott CL, Baerwald MR, Darling J, Goodwin KD, Meyer RS, Timmers MA, Thielen PM (2024) Critical considerations for communicating environmental DNA science. Environmental DNA 6(1): e472. https://doi.org/10.1002/edn3.472

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology 21(8): 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Taylor EB, Pollard S, Louie D (1999) Mitochondrial DNA variation in bull trout (*Salvelinus confluentus*) from northwestern North America: Implications for zoogeography and conservation. Molecular Ecology 8(7): 1155–1170. https://doi.org/10.1046/j.1365-294x.1999.00674.x

Thomsen PF, Kielgast J, Iversen LL, Møller PR, Rasmussen M, Willerslev E (2012) Detection of a diverse marine fish fauna using environmental DNA from seawater samples. PLOS ONE 7(8): e41732. https://doi.org/10.1371/journal.pone.0041732

Vences M, Lyra ML, Bina Perl RG, Bletz MC, Stankovic D, Lopes CM, Jarek M, Bhuju S, Geffers R, Haddad CFB, Steinfartz S (2016) Freshwater vertebrate metabarcoding on Illumina platforms using double-indexed primers of the mitochondrial 16S rRNA gene. Conservation Genetics Resources 8(3): 323–327. https://doi.org/10.1007/s12686-016-0550-y

Wang S, Yan Z, Hanfling B, Zheng X, Wang P, Fan J, Li J (2021) Methodology of fish eDNA and its applications in ecology and environment. The Science of the Total Environment 755: 142622. https://doi.org/10.1016/j.scitotenv.2020.142622

Wangensteen OS, Palacín C, Guardiola M, Turon X (2018) DNA metabarcoding of littoral hard-bottom communities: High diversity and database gaps revealed by two molecular markers. PeerJ 6: e4705. https://doi.org/10.7717/peerj.4705

Westfall KM, Therriault TW, Abbott CL (2022) Targeted next-generation sequencing of environmental DNA improves detection of invasive European green crab (*Carcinus maenas*). Environmental DNA 4(2): 440–452. https://doi.org/10.1002/edn3.261

Zaiko A, Pochon X, Garcia-Vazquez E, Olenin S, Wood SA (2018) Advantages and limitations of environmental DNA/RNA tools for marine biosecurity: Management and

surveillance of non-indigenous species. Frontiers in Marine Science 5: 322. https://doi.org/10.3389/fmars.2018.00322

Zaiko A, Greenfield P, Abbott C, von Ammon U, Bilewitch J, Bunce M, Cristescu ME, Chariton A, Dowle E, Geller J, Ardura Gutierrez A, Hajibabaei M, Haggard E, Inglis GJ, Lavery SD, Samuiloviene A, Simpson T, Stat M, Stephenson S, Sutherland J, Thakur V, Westfall K, Wood SA, Wright M, Zhang G, Pochon X (2022) Towards reproducible metabarcoding data: Lessons from an international cross-laboratory experiment. Molecular Ecology Resources 22(2): 519–538. https://doi.org/10.1111/1755-0998.13485

Zinger L, Bonin A, Alsos IG, Bálint M, Bik H, Boyer F, Chariton AA, Creer S, Coissac E, Deagle BE, De Barba M, Dickie IA, Dumbrell AJ, Ficetola GF, Fierer N, Fumagalli L, Gilbert MTP, Jarman S, Jumpponen A, Kauserud H, Orlando L, Pansu J, Pawlowski J, Tedersoo L, Thomsen PF, Willerslev E, Taberlet P (2019) DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. Molecular Ecology 28(8): 1857–1862. https://doi.org/10.1111/mec.15060

Supplementary material 1

Full list of species in NAMERS with the following voucher data: collection site, collection year, institution where voucher is housed, and catalogue number of voucher

Authors: Kristen M. Westfall, Gregory A. C. Singer, Muneesh Kaushal, Scott R. Gilmore, Nicole Fahner, Mehrdad Hajibabaei, Cathryn L. Abbott

Data type: docx

Explanation note: Metadata, site information, catalogue ID.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.8.125095.suppl1

Supplementary material 2

Genbank Accession Numbers for mitogenomes and nrDNA

Authors: Kristen M. Westfall, Gregory A. C. Singer, Muneesh Kaushal, Scott R. Gilmore, Nicole Fahner, Mehrdad Hajibabaei, Cathryn L. Abbott

Data type: docx

Explanation note: Note that nrDNA grey cells are full genes, white cells with an Accession Number are partial genes, and white empty cells are missing genes. Note that N/A in the complete mitogenome column indicates it is not complete and Genbank Accession numbers for available genes for those species are in Suppl. material 3.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.8.125095.suppl2

Supplementary material 3

Genbank Accession Numbers for partial mitochondrial genes where the complete mitogenome was not recovered, blanks indicate the gene was not recovered for that species

Authors: Kristen M. Westfall, Gregory A. C. Singer, Muneesh Kaushal, Scott R. Gilmore, Nicole Fahner, Mehrdad Hajibabaei, Cathryn L. Abbott

Data type: docx

Explanation note: Nuclear ribosomal gene Accession Numbers for these species are listed in Suppl. material 2.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.8.125095.suppl3